

The Personality of Popular Facebook Users

Daniele Quercia[§], Renaud Lambiotte[‡], David Stillwell*, Michal Kosinski[†], Jon Crowcroft[§]

[§]The Computer Laboratory, University of Cambridge

[‡]Institute for Mathematical Sciences, Imperial College

*School of Psychology, University of Nottingham

[†]The Psychometrics Centre, University of Cambridge

daniele.quercia@cl.cam.ac.uk, r.lambiotte@imperial.ac.uk, david@mypersonality.org, mk583@cam.ac.uk, jon.crowcroft@cl.cam.ac.uk

ABSTRACT

We study the relationship between Facebook popularity (number of contacts) and personality traits on a large number of subjects. We test to which extent two prevalent viewpoints hold. That is, popular users (those with many social contacts) are the ones whose personality traits *either* predict many offline (real world) friends *or* predict propensity to maintain superficial relationships. We find that the predictor for number of friends in the real world (Extraversion) is also a predictor for number of Facebook contacts. We then test whether people who have many social contacts on Facebook are the ones who are able to adapt themselves to new forms of communication, present themselves in likable ways, and have propensity to maintain superficial relationships. We show that there is no statistical evidence to support such a conjecture.

Author Keywords

web 2.0, personality research, social networks

ACM Classification Keywords

J.4 Computer Applications: Social and Behavioral Sciences.

INTRODUCTION

There is no definite answer to the question of whether digital representations of ourselves on Facebook can capture much about human social relations. This question is important yet difficult to measure. Literary intellectuals and scientists alike have recently debated that question and have formed two camps. Members of the first argue that Facebook simply supplements traditional forms of communications (e.g., text messages, face-to-face contacts) and additionally helps to maintain weak and loose ties [8, 36]. This implies that people are doing online what they have been already doing offline, with the only difference that coordination and communication costs online are lower [48].

Members of the second camp argue, instead, that social-networking sites create and promote radically new forms of

communication. Jaron Lanier, for example, thinks that communication on Facebook is inherently “self-reduced” [34]. On signing up, users are asked to fill in a questionnaire in which they should specify, for instance, what they looking for; that is, whether they are looking for ‘friendship’, ‘dating’, ‘relationship’, or ‘networking’. Those are the available conditions in Facebook world, and no other condition is available. In the same vein, a large body of academic work is coming round the idea that design decisions have profound implications on people’s social capital and well-being [6, 16, 26]. For example, Burke *et al.* found that directed communication among Facebook users is associated with greater feelings of bonding social capital and lower loneliness [7]. Surprisingly, they also found that users who passively consume greater levels of content report reduced bridging and bonding social capital and increased loneliness. These findings might consequently suggest that designing a social media site that emphasizes content consumption over peer interactions would accommodate specific user needs. Worryingly, design decisions are often taken by a very small number of technologists and impact what gets broadcasted between half a billion Facebook social contacts.

There is a corollary to the two camps’ arguments: users who will become popular on Facebook are the ones who are able (have the “right” personality) to *either* attract many friends offline *or* “reduce” themselves in order to make their Facebook profiles appear more likable. This corollary reflects two extreme viewpoints but, in reality, this dichotomy might not be strict: offline and online friendship creation might well be both affected by similar factors such as being skilled at social interaction (e.g., self-presentation) or being motivated by social status. To verify the extent to which this is true, we study popularity (defined as number of social contacts) on Facebook. As we shall see in Section “Related Work”, one widely-used measure of sociometric popularity in the literature is number of friends, but this measure has been predominantly studied at a small scale. We analyze data from a Facebook application (named myPersonality) that offers its users a set of genuine personality tests and researchers access to high quality and representative samples that are comparable or better than traditional samples used in psychology research (Section “MyPersonality”). We test two competing ideas that explain creation of Facebook contacts by comparing two personality traits related to offline friend creation - Extraversion and Neuroticism - and one trait related to “reducing” oneself online - self-monitoring. In so doing, we make the following contributions:

- We study the relationship between number of contacts and personality traits that are typically associated with offline popularity, i.e., with high number of real-world friends (Section “Traditional Explanation”). We build a simple linear regression model that predicts number of contacts based on the ‘big five’ standard personality traits. Similar to what holds for offline friends, we find that also number of Facebook contacts is positively associated with Extraversion and negatively with Neuroticism. Our analysis shows that Extraversion is a predictor (albeit weak) for number of social contacts. Interestingly, despite weak individual-to-individual correlations, we find that a linear relationship between number of contacts and Extraversion is borne out extremely cleanly at population level.
- We also study the relationship between number of social contacts and a personality trait called “Self-monitoring” that predicts the ability of individuals to present themselves in likable ways and maintain superficial relationships (Section “Alternative Explanation”). We find that the relationship between number of contacts and self-monitoring is statistically negligible, if one controls for Extraversion. More specifically, based on partial correlations, one concludes that, if Extraversion is controlled for, self-monitoring does *not* explain number of contacts; by contrast, if self-monitoring is controlled for, Extraversion does explain number of contacts.
- We finally discuss theoretical implications of our findings on personality research and practical implications for viral marketing and social media security (Section “Implications”).

RELATED WORK

Social-networking sites such as Facebook not only support the maintenance of existing social ties and the formation of new ones (as Ellison *et al.* has found in the US [16] and Lewis and West in the UK [36]) but they also impact how profile owners are socially judged in the real world. Researchers have studied the factors that might influence the perception (including popularity) of a profile. The most obvious factor is the profile itself. In 2007, Kleck *et al.* found that even just the number of social contacts shown on one’s Facebook profile triggers positive social judgments in terms of popularity, pleasantness, heterosexual appeal, and confidence [32]. A year later, Tong *et al.* conducted a study among 153 US undergraduates and found that the relationship between Facebook friend count and positive social judgment is not linear (the bigger one’s count, the more positive he/she is judged to be) but quadratic: beyond a certain number (typically 300), friend count may reach a point of incredulity or foolishness [52]. By contrast, in 2010, Utz conducted a similar experiment among 124 users of the largest Dutch social-networking site and found no evidence of any linear or quadratic relationship: positive social judgment (e.g., perceived extroversion) was only weakly related to friend count. One important difference between Tong *et al.*’s study and Utz’s is that the subjects in the first study had almost twice as many social contacts as the subjects in the second.

How a profile is perceived depends not only on the profile itself but also on who views the profile. Gogolinski recently created two fictitious Facebook pages that differed from amount and type of information they displayed [18]. She showed the two profiles to 134 US undergraduates and found that the way the profiles are perceived changed depending on the viewer’s personality: lower self-monitors preferred the less detailed and more cautious profile, while higher self-monitors preferred the more detailed and generous profile. The problem is that the subjects in this study were largely women enrolled in undergraduate courses. It is not clear whether men and, more generally, individuals of different generations would hold the same view. More recently, for 167 Facebook users, Golbeck *et al.* successfully predicted these users’ five personality traits (which we will describe in Section “The Big Five”) out of their personal information and posts [19]. In Section “Discussion”, we will spell out how our work extends Golbeck *et al.*’s.

This brief literature review has offered contrasting results that are telling examples of how hard it is to generalize any finding upon a very limited and biased sample of users, largely few hundreds of undergraduates of WEIRD (White, Educated, Industrialized, Rich, and Democratic) countries [25]. Also, personality research of social-networking users has long suffered from a severe measurement problem: personality information is hard to measure, especially at scale, over time, and at the same time as unobtrusively observing user behavior.

It thus seems that a new way of measuring personality information of social-networking users is needed. Ideally, this new way should use well-established tests to record personality scores of *unbiased* and *large* samples of users whose behavioral measures are recorded *directly* as the users are on the social-networking site.

MYPERSONALITY

myPersonality is a Facebook application that users install to take a variety of genuine personality and ability tests (Figure 1). Users can also opt in and give their consent to share their personality scores and profile information, and around 40% of its 5.5 million users choose to do so. The problem with any web-based survey (including myPersonality) could be that users might respond in a careless, dishonest, or mischievous way, and that might compromise data validity [5, 46]. So it is worth emphasizing the measures that myPersonality undertakes to assure the highest quality of its databases:

1. *The right incentives for participants are in place.* As opposed to what often takes place in traditional research, subjects are not paid nor receive college credits. myPersonality users are solely motivated by the prospect of receiving reliable feedback and test results that accurately describe their personalities. Respondents can leave as many questions unanswered as they like (while researchers can easily discard tests with unanswered questions).
2. *Unreliable results are removed.* Numerous validity tests

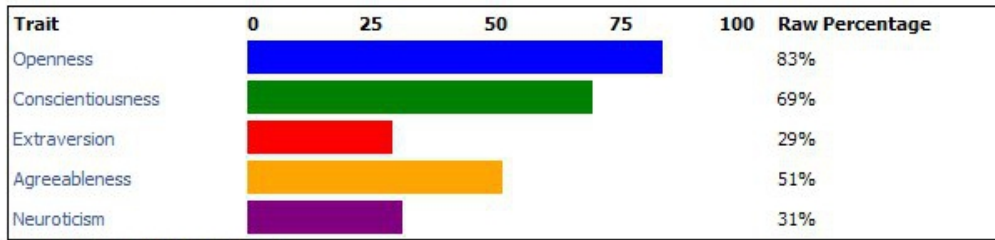


Your Personality Profile

The Big Five Personality Questionnaire & Describe Me

More Options

This will expand once your friends have described you. Click here to invite them to.



Last updated on: 08 June 2008 (Update Now)

Figure 1. Part of the myPersonality user interface.

are applied to remove the protocols that may be a product of inattentive, language incompetent, or randomly responding individuals. Consequently, the quality of the responses in our sample is at least as high as in traditional pen-and-paper studies. This is demonstrated by the scales' reliabilities which are on average higher in myPersonality samples than reported in test manuals¹. Additionally, the discriminant validity of myPersonality sample (average $r = .16$) is better than one obtained using traditional samples (average $r = .20$ [29]).

After ensuring the right incentives are in place and removing unreliable results, myPersonality is able to obtain test results that are more reliable than those in pen-and-paper studies. Also, compared to pen-and-paper studies published in the highly-regarded Journal of Personality and Social Psychology (JPSP) in 2002 [22], myPersonality users are far less biased than those studies' subjects for gender, age, and geography. However, our dataset has its own limitations and we will later discuss those that are specific to our study (Section "Discussion").

TRADITIONAL EXPLANATION

Having this data at hand, we are able to study the relationship between personality and Facebook popularity. Previous research suggests that personality influences both offline and online activities [23]. So it is reasonable to test whether personality traits that predict popularity offline would also predict popularity online.

The Big Five

The five-factor model of personality, or the big five, consists of a comprehensive and reliable set of personality concepts [10, 20]. The idea is that an individual can be associated with five scores that correspond to five main personality traits. Table 1 lists the big five personality traits along with descriptive terms that are typically associated with them. Personality traits predict a number of real-world behaviors. They, for example, are strong predictors of how marriages

¹<http://www.mypersonality.org/wiki/>

Personality Trait	High scorers	Low scorers
Openness	Imaginative	Conventional
Conscientiousness	Organized	Spontaneous
Extraversion	Outgoing	Solitary
Agreeableness	Trusting	Competitive
Neuroticism	Prone to stress and worry	Emotionally stable

Table 1. The big five personality dimensions.

turn out: if one of the partner is high in Neuroticism, then divorce is more likely. "People high in Extraversion really do talk a lot ... When asked to think about or view something stressful and unpleasant, people high in Neuroticism really do become more upset than people low in Neuroticism. When people high in Agreeableness listen to stories, they really do pay more attention to the mental states of the characters than those low in Agreeableness" [41].

Research has consistently shown that people's scores are stable over time (they do not depend on quirks and accidents of mood) and correlate well with how others close to them (e.g., friends) see them [41].

Since the 1980s, the relationship between popularity and the big five personality traits has been widely studied. Individuals who are more outgoing and social (Extraverted) tend to report greater numbers of people in their social network [51]: they are more inclined to approach others and are more likely to receive/give social support as compared to more introverted individuals [4].

Next, we will consider the (big) five personality trait and their relationships with popularity and, more generally, with social support.

The trait of *Extraversion* is associated with descriptive terms such as sociability, activity, and excitement seeking. Individuals higher in Extraversion tend to report larger social networks, greater contact with network members, lower

Personality Trait	r
Openness	0.013
Conscientiousness	0.018
Extraversion	0.2
Agreeableness	0.039
Neuroticism	-0.068
Sex	0.012
Age	-0.154

Table 2. (Pearson) correlation coefficients between the logarithm of the number of social contacts and each of the big five personality traits, Sex and Age, as measured in the large dataset (172,952 users).

number of conflicts, and greater social status [2, 3, 24, 47, 51]

The trait of *Neuroticism* is associated with descriptive terms such as emotional lability and impulsiveness. Individuals high in Neuroticism are prone to more negative mood states (e.g., anxiety, depression), tend to withdraw from other during times of stress [35, 39, 43], and generally report less satisfaction with the support received by their social networks [30, 31].

The trait of *Agreeableness* is associated with descriptive terms such as trusting, altruistic and tender-minded. Research has shown that individuals who are high in *Agreeableness* are motivated to maintain positive relationships with others [28]. However, less is known about the relationship between *Agreeableness* and structural social support (e.g., network size, contact with network members).

The trait of *Conscientiousness* is associated with descriptive terms such as ambitious, resourceful and persistent. It is not clear the relationship between Conscientiousness and number of friends or, more generally, social support: although social support is available to those high in Conscientiousness, given their competence and resourcefulness, they may simply not require as much support from others [12, 38].

The trait of *Openness* is associated with descriptive terms such as imaginative, spontaneous, and adventurous. Openness has been found to be positively associated with network size ($r = 0.23$) [50]. More generally, studies about the relationship between Openness and social support have been inconsistent.

To recap, the trait of Extraversion has been found to be associated with number of real-world friends and with structural measures of social support, including network size and contact with network members. Instead, Neuroticism has been associated with negative social interactions. Finally, the findings for the traits of Agreeableness, Conscientiousness and Openness are inconsistent.

Results

We take a sample of 172,952 Facebook users who live in the United States² and have taken the big five personality test,

²In Section “Discussion”, we detail why we have chosen users from a single country.

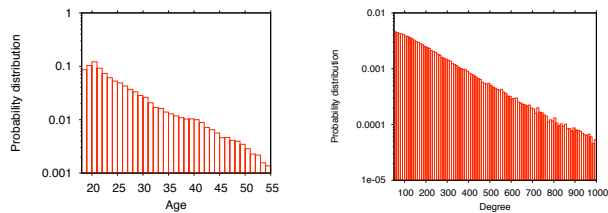


Figure 2. Probability distributions of age (*left*) and number of friends (*right*).

Variables	β	t-test	β	t-test
Openness	-0.031	-12.6	-	-
Conscientiousness	0.004	1.5	-	-
Extraversion	0.174	79.7	0.171	79.8
Agreeableness	0.013	5.1	-	-
Neuroticism	0.005	2	0	0
Sex	0.005	1.4	-	-
Age	-0.015	-64.5	-0.014	-64.7

Table 3. Prediction model for number of friends. (*left*) A linear regression model between $\log(k)$ and the big five personality scores, plus Sex and Age. (*right*) A linear regression model between $\log(k)$ and Extraversion, Neuroticism and Age. The table reports t test statistics rather than p -value because, in this case, t test statistics reflect statistical significance of variables better than p -value. In fact, in the linear regression on the left side: Openness, Extraversion, Agreeableness and Age have all p -value $< 10^{-4}$. Openness and Agreeableness are not significant for the prediction.

whose number of social contacts is between 30 and 1000 and whose age is comprised between 18 and 54. This group is composed of 103,482 women (60%) and 69,470 men (40%) with a median age of 23. Figure 2 plots their age and degree distributions. Our analysis unfolds in three steps:

1. *Determine relevant predictors for number of Facebook contacts.* We study the relationship between personality traits and the logarithm $\log(k)$ of the number of social contacts. We are interested in the logarithm rather than the number of social contacts simply because of the large variability of number of contacts (right graph in Figure 2). The five traits are all normally distributed, suggesting the lack of selection bias among our subjects. We study the Pearson product-moment correlation between $\log(k)$ and each of the (big) five personality traits, plus two additional attributes, namely sex and age. Pearson’s correlation $r \in [-1, 1]$ is a measure of the linear relationship between two random variables. Table 2 summarizes the results. Weak correlations are found with Extraversion (0.2), as expected, and with Age (-0.154). The latter indicates that older individuals tend to have significantly fewer Facebook contacts than younger ones. Among the remaining personality traits, Neuroticism (-0.068) has a very weak negative correlation with $\log(k)$. These correlations and corresponding magnitudes are consistent with previous research about both offline and online popularity.
2. *Determine the statistical significance of the predictors.* Since Pearson correlation coefficients are low, it might be

worth checking the statistical significance of each predictor. To this end, we build a regression model that predicts $\log(k)$ based on the previous variables. That is, we model the logarithm of number of friends as a linear combination of the big five personality scores, plus Sex and Age:

$$\log(k_i) = \alpha + \sum_{l=1}^5 \beta_l R_{l;i} + \beta_S S_i + \beta_A A_i + \epsilon_i$$

In the previous equation, k_i is user i 's number of contacts, R_i is a vector that contains user i ' big five personality scores, S_i is i 's Sex (0 if female or 1 if male), A_i i 's Age and ϵ_i is the error term. The extent to which the regression predicts $\log(k_i)$ is reflected in a measure called R^2 - the higher R^2 , the better the fit of the model. In this case, $R^2 = 0.064$, which is small. However, to understand the statistical significance of each predictor, one examines the predictor's t statistic - the higher in absolute value, the more significant the predictor. From Table 3, we learn that Extraversion and Age have the highest t statistics. This implies that, to model the logarithm of number of Facebook contacts, Extraversion and Age are sufficient. This hypothesis is verified by checking that the addition of any other variable in the regression (e.g., Neuroticism, Openness) does not improve results (R^2 remains roughly the same).

3. *Determine the nature of the linear relationships of predictors.* We now know that Extraversion is the personality trait that correlates with the logarithm of Facebook contacts. However, since R^2 (the model's predictive power) is small, it is not clear yet to which extent the relationship between the logarithm and Extraversion is linear. To find this out, we segment our users in 40 bins depending on their number of social contacts (depending on $\log(k)$) and we measure the *average* personality score for each of the (big) five traits in each bin. Interestingly, despite high individual-to-individual variability reflected in our R^2 , the linear relationship between $\log(k)$ and Extraversion is borne out extremely cleanly by the aggregate data (Figure 3). Also, people at the extremes of the Extraversion scale greatly differ for number of contacts: extreme extroverts are expected to have two times as many contacts as extreme introverts.

ALTERNATIVE EXPLANATION

So far we have assumed that the meaning of friendship on Facebook is comparable to that in the real-world, but friendship on Facebook does not always have traditional connotations. Research suggests that an individual can maintain 10-20 close relationships [44] and 150 social relationships [15]. An average Facebook user has 130 friends³, which is close to the "default" size of the social group for humans [15]. However, Facebook friendships do not simply mirror offline relationships. While some Facebook friendships may represent real life strong ties, others might be more superficial links between acquaintances. Indeed, Jaron Lanier writes [34]:

"The most effective young Facebook users - the ones

³<http://www.facebook.com/press/info.php?statistics>

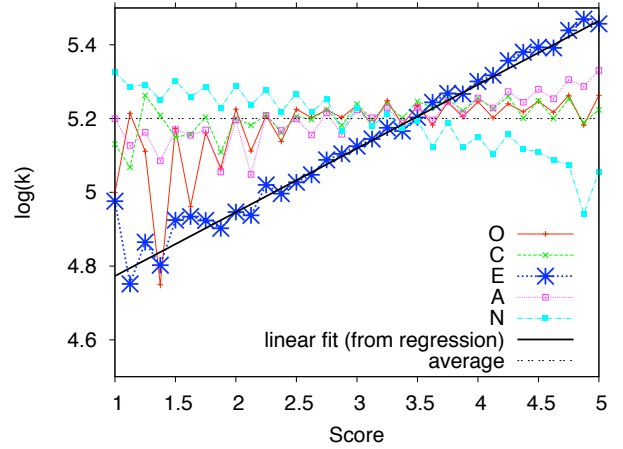


Figure 3. Average $\log(k)$ as a function of score for the 5 big personality traits. The dashed line indicates $\langle \log(k) \rangle$ across the whole population. The solid line indicates the linear fit obtained from a linear regression of $\log(k)$ with Extraversion alone.

who will probably be winners if Facebook turns out to be a model of the future they will inhabit as adults - are the ones who create successful online fictions about themselves.

They must manage offhand remarks and track candid snapshots as carefully as politician. Insincerity is rewarded, while sincerity creates lifelong taints. Certainly, some version of this principle existed in the lives of teenagers before the web came along, but not with such unyielding, clinical precision."

In Section "Related Work", we have seen that one's profile impacts one's ability to attract social contacts. Hence it is reasonable to posit that users who are able to adapt to new forms of 'self-presentation' are likely to become popular. To test whether this is true, we need a personality trait associated with the way people present themselves.

Self-monitoring

In 1974, Mark Snyder realized that personality research had failed to consider that some people's personalities are more 'fluid' than others'. Hence, he conducted in-depth interviews and found that some people are more prone to pick up on social cues and adjust how they act and how they are perceived by others. He called such individuals 'high self-monitors' [49]. High self-monitors tended to agree with statements like "I can make impromptu speeches even on topics about which I have almost no information". By contrast, low self-monitors tended to agree with statements such as "I find it hard to imitate the behavior of other people" and "I have trouble changing behavior to suit different people and different situations".

Snyder also found that high self-monitors had distinct attributes. They had unusual ability to pick up social cues and

modulate their emotional expression; to quickly learn what behaviors were appropriate in various situations; and to easily manage the perceptions of others.

Although high self-monitors modulate their behavior to fit the situation, they do not actively manipulate others to get people to like them but are actually just as genuine and principled as low self-monitors. They just pick on social cues in an unconscious effort to establish a genuine connection with others. For example, it has been shown that high self-monitors mirror their peers’ body languages without even realizing they are doing so [9].

The end result is that a high self-monitor modifies his/her behavior (and self-presentation) to fit the situation and the people in it, while a low-self monitor does not try to alter behavior very much across situations. Successful politicians, for example, tend to be high self-monitors [40].

High self-monitors modify their behavior not only offline but also online. In 2008, Lin studied the personal pages created by users of web portals [37]. He found that high self-monitors would display limited and generic information on their pages to represent themselves in likable ways, while low self-monitors would display more personal and in-depth information to accurately portray themselves. More recently, it has been shown that the same holds for Facebook [18]: high self-monitors tend to have less detailed and more cautious pages, while low self-monitors prefer more detailed pages.

Results

We take a sample of 2,165 users who live in the United States and have taken not only the big five personality test but also the self-monitoring test. Those users have a number of social contacts between 30 and 1000, with the same age interval as in the large dataset studied in the previous section. Compared to the large dataset, the correlations with the big five traits are similar (Table 4) and the regression with all variables has $R^2 = 0.041$. The only new result is that there is a weak correlation between $\log(k)$ and self-monitoring ($r = 0.089$). It follows that self-monitoring could help to predict $\log(k)$. To see the extent to which it does so, we build a linear regression with Extraversion alone. We find that the addition of self-monitoring to the regression does only marginally improve the regression’s predictions (R^2 goes from 0.037 to 0.038). That is simply because there is a strong interaction effect between self-monitoring and Extraversion ($r = 0.312$ between the two): if a regression model has Extraversion, then it would not improve its prediction by adding self-monitoring, as the two variables are strongly related. We conclude that self-monitoring does not contribute in explaining the number of Facebook contacts. This argument is also supported by the low value of self-monitoring for the t statistic (Table 4).

Also, the lack of direct dependence of $\log(k)$ on self-monitoring (SM) and its dependence on Extraversion (E) are also confirmed by partial correlations. By construction, partial correlations measure the degree of association

Variables	r	β	t-test
Openness	0.038	-0.013	-0.5
Conscientiousness	0.052	0.02	0.9
Extraversion	0.191	0.155	7.5
Agreeableness	0.04	0.01	0.4
Neuroticism	-0.069	0.005	0.3
Age	-0.045	-0.006	-2.9
Self-monitoring	0.089	-0.014	1.3

Table 4. (left) Correlation coefficients between $\log(k)$ and each of the big five personality traits, Age and self-monitoring, as measured in the small dataset (2,165 users). (right) A linear regression model between $\log(k)$ and the same variables.

between two random variables by controlling for the effect of a third variable. We compute two correlations and obtain $r(\log(k), E|SM) = 0.16$ (with $p < 10^{-5}$), and $r(\log(k), SM|E) = 0.03$ (with $p = 0.17$). The comparison of these two correlations clearly confirms our interpretation that self-monitoring is mainly correlated to $\log(k)$ due to its association to Extraversion. Finally, there is no correlation between self-monitoring and any of the four personality traits other than Extraversion.

DISCUSSION

We will now discuss the limitations of our dataset and our study, the novelty of our big five study, and the theoretical and practical implications.

Limitations of personality data. All users of myPersonality choose to participate motivated primarily by the prospect of taking and receiving feedback from a high quality personality questionnaire used in actual scientific research. Although this encourages them to respond to the questions accurately and honestly, it means that the sample is a convenience sample of self-selected users interested in their personality, and so the results may not represent the Facebook population as a whole. For example, the number of contacts for the average myPersonality user is 194 whereas Facebook reports an average of 130⁴. This is most likely because number of Facebook contacts is confounded by Facebook activity, and active Facebook users are more likely to use applications. In future, Facebook activity should be controlled for, although this information is not available automatically from Facebook’s API. Nevertheless, there are reasons to be confident in our conclusions. First, the demographic and geographic distribution of myPersonality users is closer to that of the general population than that reported in social psychology studies which typically rely on undergraduates [22]. Second, our results concentrate on the relationships between personality and number of Facebook contacts, so any sampling bias would have to affect this relationship, and it is unclear why application users would have a different association between personality and number of contacts than the general Facebook population.

⁴<http://www.facebook.com/press/info.php?statistics>

Limitation of our study. The only variable under study has been number of contacts, and that is because it is the only unbiased network metric readily-available from our application. Ideally, we would have included other variables of interest, but those variables either have not been captured because of technical limitations imposed by Facebook (e.g., frequency of interaction among users) or are biased if taken at face value (e.g., since we do not have information about non-myPersonality users, one’s network clustering coefficient could be computed, but only upon a biased network that contains myPersonality users and excludes non-myPersonality ones). We are currently exploring techniques that would allow us to reduce such biases or, at least, quantify them.

Novelty of the Big Five Study. In section “Traditional Explanation”, we have analyzed the results for the Big Five test, and Golbeck *et al.* have already done so [19]. We build upon their work in that we use the same (data analysis) methodology and extend previous experimental results as follows:

- The results in Golbeck *et al.*’s work are about 167 participants, 82.6% (138) of whom are from the United States and the remaining subjects come from India (8), Australia (7), Italy (7), and others (7). Our results are about 172,952 subjects in United States. We chose subjects from a single country for two main reasons. First, our test is the commonly-used US version of the Big Five test. Second, the Big Five factors were found to function differently across samples from different countries: although the same traits may exist in different cultures, there is a large bias at the metric and scalar levels of analysis across cultures [42]. Therefore, mean-level cross-cultural comparisons should be done only after ensuring measurement equivalence. To interpret our results correctly, we have users from the United States - the country from which we had the most representative sample.
- In our study, all the five personality traits have statistically significant (albeit weak) correlations. Furthermore, age is an additional factor that shows a statistically significant correlation and its magnitude reflects a weak yet important correlation as it is comparable to that with Extraversion. Also, the correlation is negative (i.e., younger people tend to have more social contacts on Facebook). By contrast, the opposite holds in Twitter: we have studied the personality of Twitter users and have found a positive correlation between number of followers/following and age [45]. These contrasting results reflect the different user demographics in the two platforms and reflects the importance of controlling for age.
- In line with previous work, our individual-to-individual correlations are all weak. Yet, at population-level, we report, for the first time, a clear linear relationship between number of contacts and Extraversion (Figure 3).

Theoretical Implications. It is important to study the nature of Facebook relations, not least because they do not stay

confined online but do impact access to real-world opportunities. They have been shown, for example, to: (1) *Increase social capital*: A large number of weak social ties on Facebook becomes a source of social capital [16]; (2) *Trigger positive social judgments in the real world*: Just the number of social contacts on one’s Facebook profile triggers positive social judgments in terms of popularity, pleasantness, heterosexual appeal, and confidence [32].

By studying the self-monitoring trait and the big five personality traits, we have tried to understand the nature of Facebook relationships and have asked whether they are either superficial (typical of high self-monitoring individuals) or “real” (individuals who are high in Extraversion and low in Neuroticism typically have a large number of offline friends). As our Introduction has exemplified, the dichotomy real vs. superficial has been widely debated, but we have now partially quantified it. Extraversion is not only weakly correlated with number of online social contacts (as it is with number of offline friends), but its relationship with number of contacts is also strongly linear (Figure 3). By contrast, self-monitoring, which correlates with superficial relationships in the real world, does not have any explanatory power for number of contacts.

Practical Implications. Past research has shown that, from textual content accessible (e.g., blog posts, Facebook pages dedicated to brands), one could infer personality scores [19, 21, 33, 54]. We have shown that, from those scores, one is then able to partially predict the number of social contacts on Facebook based on the three variables Extraversion, Age, and Neuroticism. If we then build a simple Naive Bayesian model with those three variables, we find that we are able to predict whether Facebook users are likely to be social hubs or poorly-connected. More specifically, if we define all user with number of social contacts in the range [500, 1000] to be social hubs and perform a 10-fold cross-validation on the big dataset (that of Section “Traditional Explanation”), we obtain that 83.68% of the social hubs are correctly classified by the model. The ability to identify social hubs may benefit a variety of individuals such as:

Social Media Marketers. Those individuals are interested in identifying social hubs and poorly-connected users in order to:

- Launch marketing campaigns that, to be successful, either target few social hubs [14] or many poorly-connected individuals who are passionate about certain brands and have spare time at hand [13].
- Measure brand awareness. Marketing agencies often measure, for a given brand, the quality of a community around it and they do so based on the presence of social hubs in that community [1].

Phishing Attackers. Those individuals are interested in avoiding social hubs and finding poorly-connected users who tend to be more gullible and less technological-savvy than

typical Internet users [27].

Those examples not only point to the use of public data for questionable (marketing) reasons but also inform two research areas other than personality research. First, they inform the privacy literature in that our results might raise awareness among online users about the privacy-sensitive predictions that are possible from publicly-available data. This is similar to Crandall *et al.*'s very recent work, which showed that, from publicly available geo-referenced Flickr pictures, one is able to infer several coincidences (eg, two people taking picture at the same place at the same time) that in turn reveal people's personal social contacts [11]. More importantly, in addition to raise awareness around privacy issues, our findings might well inform the design of new privacy protection mechanisms. Having the necessary means to identify the social-networking users who are likely to be targeted by viral campaigns or phishing attacks, one could design feedback mechanisms for those targets. That is important not only for underage targets (e.g., for designing child protection mechanisms) but for targets of any demographic. That is largely because, although feedback on "who views what" has been shown to be a central element for helping people manage their privacy, social-networking sites currently provide very little of it [53].

The second area is social network modeling. Our research suggests that people are not alike in that they have different tendencies of attracting attention/creating links and, consequently, adding hidden properties such as personality traits could be crucial for modeling online social networks in realistic ways.

CONCLUSION

Creators of social-networking sites make specific design choices and those choices may impact the way people communicate on those sites. Facebook, for example, allows its users to 'poke' social contacts or to exchange personal trivia. However, based on that, one cannot conclude that Facebook is surely "reconfiguring" our ways of communicating with each other or that "the most effective young Facebook users ... are the ones who create successful online fictions about themselves" [34]. Those are seemingly reasonable speculations for which, unfortunately, we found little evidence. Popular Facebook users tend to have the same personality as popular people in the real world, suggesting that the nature of online interactions does not significantly differ from that of real world interactions. So, if design choices were made by Facebook creators, then those choices had little impact on which personality traits determine popularity. We could speculate that it is more likely that Facebook users simply devise their own patterns and structures within status/comment free-form fields outside the narrow creators' original straitjackets. If that is the case, to explain interactions on social-networking sites, the words of William Gibson might have been more apt: "The street finds its own uses for things" [17].

Acknowledgments. We thank EPSRC for its financial support. This work was in part funded by RCUK through the

Horizon Digital Economy Research grant (EP/G065802/1). We also thank Paul Expert for his assistance in the statistical analysis. We finally thank the three anonymous reviewers.

REFERENCES

1. R. Algesheimer, U. Dholakia, and A. Herrmann. The Social Influence of Brand Community: Evidence from European Car Clubs. *Journal of Marketing*, 69, July 2005.
2. C. Anderson, O. John, D. Keltner, and A. Kring. Who attains social status? Effects of personality and physical attractiveness in social groups. *Journal of Personality and Social Psychology*, 81, July 2001.
3. D. S. Berry, J. K. Willingham, and C. A. Thayer. Affect and personality as predictors of conflict and closeness in young adults' friendships. *Journal of Research in Personality*, 34, March 2000.
4. N. Bowling, T. Beehr, and W. Swader. Giving and receiving social support at work: the roles of personality and reciprocity. *Journal of Vocational Behaviour*, 67, December 2005.
5. T. Buchanan and J. Smith. Using the Internet for psychological research: Personality testing on the World Wide Web. *British Journal of Psychology*, 90, March 2002.
6. M. Burke, R. Kraut, and C. Marlow. Social Capital on Facebook: Differentiating Uses and Users. In *Proceedings of the 29th ACM Conference on Human Factors in Computing Systems (CHI)*, May 2011.
7. M. Burke, C. Marlow, and T. Lento. Social Network Activity and Social Well-Being. In *Proceedings of the 28th ACM Conference on Human Factors in Computing Systems (CHI)*, April 2010.
8. J. T. Cacioppo and W. Patrick. *Loneliness: Human Nature and the Need for Social Connection*. W. W. Norton & Company, 2008.
9. C. M. Cheng and T. Chartrand. Self-Monitoring Without Awareness: Using Mimicry as a Nonconscious Affiliation Strategy. *Journal of Personality and Social Psychology*, December 2003.
10. P. Costa and R. McCrae. *The Revised NEO Personality Inventory (NEO-PI-R)*. SAGE Publications, 2005.
11. D. Crandall, L. Backstrom, D. Cosley, S. Suri, D. Huttenlocher, and J. Kleinberg. Inferring social ties from geographic coincidences. *Proceedings of the National Academy of Sciences*, 107, December 2010.
12. C. Dehle and J. Landers. You can't always get what you want, but can you get what you need? Personality traits and social support in marriage. *Journal of Social and Clinical Psychology*, 24, November 2005.

13. P. S. Dodds, R. Muhamad, and D. J. Watts. An Experimental Study of Search in Global Social Networks. *Science*, 301:827–829, August 2003.
14. W. Duan, Z. Chen, Z. Liu, and W. Jin. Efficient target strategies for contagion in scale-free networks. *Physical Review E*, 72:026133, August 2005.
15. R. Dunbar. Coevolution of neocortical size, group size and language in humans. *Behavioral and Brain Sciences*, 16, 1993.
16. N. Ellison, C. Steinfield, and C. Lampe. The benefits of Facebook “friends”: Exploring the relationship between college students’ use of online social networks and social capital. *Journal of Computer-Mediated Communication*, 12, July 2007.
17. W. Gibson. *Burning Chrome*. Ace Books, 1987.
18. T. B. Gogolinski. Effects Of Self-Monitoring and Public Self-Consciousness on Perceptions of Facebook Profiles. *Colonial Academic Alliance Undergraduate Research Journal*, 1, 2010.
19. J. Golbeck, C. Robles, and K. Turner. Predicting personality with social media. In *Proceedings of the 29th ACM Conference on Human Factors in Computing Systems (CHI)*, May 2011.
20. L. Goldberg, J. Johnson, H. Eber, R. Hogan, M. Ashton, R. Cloninger, and H. Gough. The international personality item pool and the future of public-domain personality measures. *Journal of Research in Personality*, 40, February 2006.
21. S. Gosling and S. Vazire. Personality impressions based on facebook profiles. In *Proceedings of the International Conference on Weblogs and Social Media (ICWSM)*, March 2007.
22. S. Gosling, S. Vazire, S. Srivastava, and O. John. Should we trust web-based studies? A comparative analysis of six preconceptions about Internet questionnaires. *American Psychologist*, 59, February 2004.
23. E. Hargittai. Whose space? Differences among users and non-users of social network sites. *Journal of Computer-Mediated Communication*, 13, 2008.
24. S. Henderson. *Neurosis and the social environment*. New York: Academic Press, 1981.
25. J. Henrich, S. Heine, and A. Norenzayan. The weirdest people in the world? *Behavioral and Brain Sciences*, 33, June 2010.
26. B. Hogan. The Presentation of Self in the Age of Social Media: Distinguishing Performances and Exhibitions Online. *Bulletin of Science Technology Society*, December 2010.
27. T. Jagatic, N. Johnson, A. Nathaniel, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, 50, October 2007.
28. L. Jensen-Campbell and W. Graziano. Agreeableness as a moderator of interpersonal conflict. *Journal of Personality*, 71, January 2004.
29. O. P. John and S. Srivastava. The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In *Handbook of Personality: Theory and Research*. Guilford Press, 1999.
30. G. D. Jong, E. V. Sonderen, and P. Emmelkamp. A comprehensive model of stress: the roles of experience stress and Neuroticism in explaining the stress- distress relationship. *Psychotherapy and Psychosomatics*, 68, 1999.
31. B. R. Karney and T. N. Bradbury. The longitudinal course of marital quality and stability: a review of theory, methods and research. *Psychological Bulletin*, 118, July 1995.
32. C. A. Kleck, C. A. Reese, D. Z. Behnken, and S. S. Sundar. The company you keep and the image you project: Putting your best face forward in online social networks. In *Proceedings of the International Communication Association*, May 2007.
33. A. Kramer, D. Adam, and K. Rodden. Word usage and posting behaviors: modeling blogs with unobtrusive data collection methods. In *Proceeding of the 26th Annual Conference on Human Factors in Computing Systems (CHI)*, April 2008.
34. J. Lanier. *You Are Not a Gadget: A Manifesto*. Knopf, 2010.
35. D. Lee-Baggley, M. Preece, and A. DeLongis. Coping with interpersonal stress: role of Big Five traits. *Journal of Personality*, 73, July 2005.
36. J. Lewis and A. West. ‘Friending’: London-based undergraduates’ experience of Facebook. *New media and Society*, 11, September 2009.
37. C. S. Lin. Exploring the Personality Trait of Self-Monitoring on Technology Usage of Web Portals. *CyberPsychology and Behavior*, 11, April 2008.
38. L. Lu and M. Argyle. Receiving and giving support: effects on relationships and well-being. *Psychology Quarterly*, 5, April 1992.
39. R. McCrae and P. Costa. Personality, coping, and coping effectiveness in an adult sample. *Journal of Personality*, 54, 1986.
40. A. Mehra, M. Kilduff, and D. Brass. The Social Networks of High and Low Self-monitors: Implications for Workplace Performance. *Administrative Science Quarterly*, 11, March 2000.
41. D. Nettle. *Personality: What Makes You the Way You Are*. Oxford University Press, 2007.
42. C. D. Nyea, B. W. Roberts, G. Saucier, and X. Zhou. Testing the measurement equivalence of personality adjective items across cultures. *Journal of Research in Personality*, July 2008.

43. T. B. O'Brien and A. DeLongis. The interactional context of problem-, emotion-, and relationship-focused coping: the role of the Big Five personality factors. *Journal of Personality*, 64, December 1996.
44. M. Parks. *Personal networks and personal relationships*. Routledge, 2006.
45. D. Quercia, M. Kosinski, D. Stillwell, and J. Crowcroft. Our Twitter Profiles, Our Selves: Predicting Personality with Twitter. In *Proceedings of the 3rd IEEE Conference on Social Computing (SocialCom)*, October 2011.
46. U. Reips. *The Web Experiment Method: Advantages, disadvantages, and solutions*. CA Academic Press, 2000.
47. D. Russell, B. Booth, D. Reed, and P. Laughlin. Personality, social networks, and perceived social support among alcoholics: a structural equation analysis. *Journal of Personality*, 65, September 1997.
48. C. Shirky. *Here Comes Everybody*. Allen Lane, 2008.
49. M. Snyder. Self-monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 30, 1974.
50. R. Swickert. Personality and social support processes. *The Cambridge Handbook of Personality Psychology*, 2009.
51. R. Swickert, C. Rosentreter, J. Hittner, and J. Mushrush. Extraversion, social support processes, and stress. *Personality and Individual Differences*, 32, April 2002.
52. S. T. Tong, B. V. D. Heide, L. Langwell, and J. B. Walther. Too Much of a Good Thing? The Relationship Between Number of Friends and Interpersonal Impressions on Facebook. *Journal of Computer-Mediated Communication*, April 2008.
53. J. Y. Tsai, P. Kelley, P. Drielsma, L. F. Cranor, J. Hong, and N. Sadeh. Who's viewed you?: the impact of feedback in a mobile location-sharing application. In *Proceedings of the 27th International Conference on Human factors in Computing Systems (CHI)*, 2009.
54. S. Vazire and S. Gosling. e-Perceptions: Personality Impressions Based on Personal Websites. *Journal of Personality and Social Psychology*, 87, July 2004.